View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

SCIENCE DIRECT

Theoretical Computer Science 352 (2006) 306–317

Theoretical Computer Science

www.elsevier.com/locate/tcs

# Combinatorial properties of smooth infinite words☆

S. Brlek[a,*], S. Dulucq[b], A. Ladouceur[a], L. Vuillon[c]

[a] *LaCIM, Dept. Informatique, Université du Québec à Montréal, C. P. 8888 Succursale "Centre-Ville", Montréal, Qué., Canada H3C 3P8*
[b] *LABRI, Université Bordeaux I, 351, Cours de la Libération, F33405 Talence Cedex, France*
[c] *LAMA, UMR 5127 CNRS, Université de Savoie, 73376 Le Bourget du Lac, France*

Received 6 December 2005; accepted 12 December 2005

Communicated by M. Crochemore

**Abstract**

We describe some combinatorial properties of an intriguing class of infinite words, called *smooth*, connected with the Kolakoski one, $K$, defined as the fixed point of the run-length encoding $\Delta$. It is based on a bijection on the free monoid over $\Sigma = \{1, 2\}$, that shows some surprising mixing properties. All words contain the same finite number of square factors, and consequently they are cube-free. This suggests that they have the same complexity as confirmed by extensive computations. We further investigate the occurrences of palindromic subwords. Finally, we show that there exist smooth words obtained as fixed points of substitutions (realized by transducers) as in the case of $K$.
© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Combinatorics; Smooth words; Kolakoski word; Substitutions

## 1. Introduction

The classification of infinite words over a finite alphabet by using properties like avoidance of some patterns, or existence of some others, is one of the problems considered by Axel Thue in a series of papers [18,19], for which Berstel [1] provided an annotated translation. The pioneering work of Thue on overlap-free and square-free words led to the discovery of infinite classes of words on a finite alphabet sharing these properties. In these classes, the infinite Thue–Morse word

$$M = \lim_{n\to\infty} \mu^n(1) = 1221211221121221\cdots$$

obtained by iteration of the morphism defined over the two-letter alphabet $\Sigma = \{1, 2\}$ by $\mu(1) = 12$; $\mu(2) = 21$, is an infinite overlap-free word, which is characteristic of its class. Among the (popular) patterns, palindromes play an important role and, precisely, they are essential in order to construct infinite overlap-free words [19]. Moreover, infinite overlap-free words are characterized by means of the morphism $\mu$ and are recurrent, that is, they have the property that every factor appears infinitely many often.

☆ With the support of NSERC (Canada).
* Corresponding author.
*E-mail addresses:* brlek@lacim.uqam.ca (S. Brlek), Serge.Dulucq@labri.fr (S. Dulucq), Laurent.Vuillon@univ-savoie.fr (L. Vuillon).

0304-3975/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2005.12.003

Thue's work remained forgotten for a while and some of its results were rediscovered by Morse [16] who introduced several complexity measures among which the number $P(n)$ of different factors of each length. He also characterized the class of Sturmian words by the property $P(n) = n + 1$. Other characterizations were provided more recently, and especially that of de Luca and Mignosi (see [15]) which is based on palindromic factorizations. Again, the Sturmian words are recurrent, and among them lives the Fibonacci word

$$\phi^{\omega}(1) = 12112121121121211211211212112112\cdots$$

obtained by iterating the morphism $\phi$ defined by $\phi(1) = 12$; $\phi(2) = 1$. In this paper we describe a general framework for the study of another particular class of infinite words over the 2-letter alphabet $\Sigma = \{1, 2\}$. This class is invariant under the action of the run-length encoding operator, and is related to the curious Kolakoski word

$$K = 22112122122112112212112122112112122122112122121121122\cdots$$

which attracted considerable attention by showing some intriguing combinatorial properties, consisting mainly of a set of conjectures due to Dekking [7]. In particular it is not known whether $K$ is recurrent or not, if the set of its factors is closed under permutation of letters or mirror image, if the density of 1's is equal to $\frac{1}{2}$. The (finite) palindromes of the elements of this class are characterized by means of the palindromic closure of the prefixes of the Kolakoski word and reveal an interesting perspective for understanding some of the conjectures [3]. In particular, recurrence, mirror invariance and permutation invariance are all direct consequences of the presence in $K$ of these palindromes. This last assumption however remains a conjecture.

Other regularities such as squares, overlaps can be studied in this framework and extend the work of Carpi [4]. This work is an excerpt/extension of the Master thesis of Annie Ladouceur [11], which also contains numerous computations performed with an efficient library of functions, and these computations enabled us to discover the combinatorial properties presented here.

## 2. Definitions and notation

Let us consider a finite *alphabet* of *letters* $\Sigma$. A *word* is a finite sequence of letters $w : [1..n] \longrightarrow \Sigma\ , \quad n \in \mathbb{N}$, of length $n$, and $w[i]$ denotes its $i$th letter. The set of $n$-length words over $\Sigma$ is denoted $\Sigma^n$. By convention, the *empty* word is denoted $\varepsilon$ and its length is 0. The free monoid generated by $\Sigma$ is defined by $\Sigma^* = \bigcup_{n\geqslant 0}\Sigma^n$. The set of right infinite words is denoted by $\Sigma^{\omega}$ and $\Sigma^{\infty} = \Sigma^* \cup \Sigma^{\omega}$. Adopting a consistent notation for sequences of positive integers, $\mathbb{N}^* = \bigcup_{n\geqslant 0}\mathbb{N}^n$ is the set of finite sequences and $\mathbb{N}^{\omega}$ is that of infinite ones. Given a word $w \in \Sigma^*$, a *factor* $f$ of $w$ is a word $f \in \Sigma^*$ satisfying

$$\exists x, y \in \Sigma^*, \quad w = xfy.$$

If $x = \varepsilon$ (resp. $y = \varepsilon$) then $f$ is called *prefix* (resp. *suffix*). The set of all factors of $w$ is denoted by $F(w)$, and the set of those of length $n$ is $F_n(w) = F(w) \cap \Sigma^n$. Finally, $\mathrm{Pref}(w)$ denotes the set of all prefixes of $w$. The length of a word $w$ is $|w|$, and the number of occurrences of a factor $f \in \Sigma^*$ is $|w|_f$. Clearly, the length of a word is given by the number of its letters,

$$|w| = \sum_{\alpha\in\Sigma} |w|_{\alpha}. \tag{1}$$

A *block* of length $k$ is a factor of the particular form $f = \alpha^k$, with $\alpha \in \Sigma$. If $w = pu$, and $|w| = n$, $|p| = k$, then $p^{-1}w = w[k+1]\ldots w[n] = u$ is the word obtained by erasing $p$. As a special case, when $|p| = 1$ we obtain the *shift* function defined by $s(w) = w[2]\ldots w[n]$. The mirror image $\widetilde{u}$ of $u \in \Sigma^n$ is the unique word satisfying

$$u[i] = u[n - i + 1] \quad \forall\ 1\leqslant i\leqslant n.$$

A *palindrome* is a word $p$ such that $p = \widetilde{p}$. A factor of the form $uu$ is called a *square*, and an *overlap* is a factor of the form $xuxux$, where $x$ is a nonempty factor. For a language $L \subseteq \Sigma^{\infty}$, we denote by $\mathrm{Pal}(L)$, $\mathrm{Squares}(L)$, $\mathrm{Overlaps}(L)$ the sets, respectively, of its palindromes, square and overlapping finite factors. Over the restricted alphabet $\Sigma = \{1, 2\}$,

there is a usual length preserving morphism, the swapping of the letters, defined by $\overline{1} = 2$ and $\overline{2} = 1$, which extends to words as follows. The complement of $u \in \Sigma^n$, is the word

$$\overline{u} = \overline{u[1]}\,\overline{u[2]}\,\overline{u[3]}\cdots\overline{u[n]}\,.$$

Note that, the complement corresponds to a *permutation* of letters of $\Sigma$. The occurrences of factors play an important role and an infinite word $w$ is recurrent if it satisfies the condition

$$u \in F(w) \Longrightarrow |w|_u = \infty.$$

Clearly, every periodic word is recurrent, and there exist recurrent but nonperiodic words, the Thue–Morse word $M$ being one of these [16]. Finally, two words $u$ and $v$ are *conjugate* when there are words $x, y$ such that $u = xy$ and $v = yx$. The conjugacy class of a word $u$ is denoted by $[u]$, and the length is invariant under conjugacy so that it makes sense to define $|[u]| = |u|$.

## 3. Run-length encoding

The widely known *run-length encoding* is used in many applications as a method for compressing data. For instance, the first step in the algorithm used for compressing the data transmitted by Fax machines, consists of a run-length encoding of each line of pixels. It also was used for the enumeration of factors in the Thue–Morse sequence [2].

Let $\Sigma = \{1, 2, \}$ be an ordered alphabet. Then every nonempty word $w \in \Sigma^+$ can be uniquely written as a product of factors as follows:

$$w = \begin{cases} 1^{i_1}2^{i_2}1^{i_3}\cdots & \text{if } w \in 1\cdot\Sigma^*, \\ 2^{i_1}1^{i_2}2^{i_3}\cdots & \text{if } w \in 2\cdot\Sigma^*, \end{cases}$$

where $i_j > 0$. The operator giving the size of the blocks appearing in the coding is a function $\Delta : \Sigma^* \longrightarrow \mathbb{N}^*$,

$$\Delta(w) = [i_1, i_2, \cdots, i_k],$$

with the convention $\Delta(\varepsilon) = 0$, which is easily extended to infinite words as $\Delta : \Sigma^\omega \longrightarrow \mathbb{N}^\omega$.

**Example.** Let $w = 12212211$, then $w = 1^1 2^2 1^1 2^2 1^2$, and $\Delta(w) = [1, 2, 1, 2, 2]$. Often the punctuation and the parentheses are omitted in order to manipulate the more compact notation $\Delta(w) = 12122$.

This example is particular: indeed, the coding integers coincide with the alphabet on which is written $w$, so that $\Delta$ can be viewed as a partial function $\Delta : \{1, 2\}^* \longrightarrow \{1, 2\}^*$. Although a general theory can be done over arbitrary alphabets, with the manner of Lamas [12], we restrict from here on the study to this case, i.e. to words over the two-letter alphabet $\Sigma = \{1, 2\}$ and not having 111 or 222 as factors.

The function $\Delta$ is a contraction, that is, for every word $w \in \Sigma^*$ we have

$$|\Delta(w)| \leqslant |w|, \tag{2}$$

and equality holds when

$$w \in \{\varepsilon, 2\}\cdot(12)^*\cdot\{\varepsilon, 1\}. \tag{3}$$

The function $\Delta$ is not bijective because $\Delta(w) = \Delta(\overline{w})$. However, pseudo-inverse functions

$$\Delta_1^{-1}, \Delta_2^{-1} : \Sigma^* \longrightarrow \Sigma^*$$

can be defined by

$$\Delta_\alpha^{-1}(u) = \alpha^{u[1]}\overline{\alpha}^{u[2]}\alpha^{u[3]}\overline{\alpha}^{u[4]}\cdots \quad \text{for } \alpha \in \{1, 2\}, \tag{4}$$

and $\forall u \in \Sigma^*$, we have $\Delta_2^{-1}(u) = \overline{\Delta_1^{-1}(u)}$. For later use, given a word $x \in \Sigma^n$ of length at least 2, we define $\Delta_x^{-n}$ by $\Delta_x^{-n}(u) = \Delta_{x[1]}^{-1}(\Delta_{s(x)}^{-n+1}(u))$. The operator $\Delta$ can be iterated. Since $\Delta(1) = 1$, arriving at a word of length 1 provides no impediment to iterating the operator.

**Example.** Let $w = 12211211$. The successive application of $\Delta$ gives:
$\Delta^0(w) = 12211211$;
$\Delta^1(w) = 12212$;
$\Delta^2(w) = 1211$;
$\Delta^3(w) = 112$;
$\Delta^4(w) = 21$;
$\Delta^5(w) = 11$;
$\Delta^6(w) = 2$.

Looking at the column word

$$u = \Delta^0(w)[1] \cdots \Delta^6(w)[1] = 1111212,$$

the initial word $w$ can be retrieved, starting from the bottom and writing the prescribed number of consecutive letters. Using the notation above, we have

$$w = \Delta_{111121}^{-6}(2).$$

A natural question concerns the reversibility of this construction. The fact that

$$\Delta(1) = \Delta(2) = 1 = \Delta^k(1) \quad \forall k \in \mathbb{N}$$

shows that the column word $u' = 11112121^k$, $\forall k \geqslant 0$ also permits the retrieval of $w$. To avoid this redundancy, it suffices to restrict the column words ending with a 2. Moreover, in order to keep the coding alphabet constant $\Sigma = \{1, 2\}$, we define the set

$$\Delta_\Sigma^k = \{w \in \Sigma^+ \mid (\Delta^k(w) = 2) \wedge (\forall j, 1 \leqslant j \leqslant k-1, \Delta^j(w) \in \Sigma^+)\},$$

and denote $\Delta_\Sigma^+ = \bigcup_{k \geqslant 1} \Delta_\Sigma^k$. Therefore, the desired representation is

$$\Phi : \Delta_\Sigma^+ \longrightarrow \Sigma^+,$$
$$\Phi(w)[j+1] = \Delta^j(w)[1] \quad \text{for } 0 \leqslant j \leqslant k. \tag{5}$$

Consequently, the inverse of $\Phi$ is defined as follows. Let $u \in \Sigma^n$, $n > 0$, then

$$\Phi^{-1}(u) = \Delta_{u[1..n-1]}^{n-1}(u[n]), \tag{6}$$

or inductively by $\Phi^{-1}(u) = w_1$, where

$$w_n = u[n],$$
$$w_j = \Delta_{u[j]}^{-1}(w_{j+1}) \quad \forall j \text{ such that } 1 \leqslant j < n.$$

Of course, this bijection extends to infinite words, provided some precautions are taken.

**Definition 1.** An infinite word $W \in \Sigma^\omega$ is said to be *smooth* if and only if $\forall k \in \mathbb{N}$, $\Delta^k(W) \in \Sigma^\omega$.

Let $\mathcal{K}$ denote the set of all infinite smooth words. The elements of the set $F(\mathcal{K})$ of finite subwords of $\mathcal{K}$ are also called *smooth*. The extension is $\Phi : \mathcal{K} \longrightarrow \Sigma^\omega$, denoted and defined identically by (5).

The operator $\Delta$ has two fixpoints, that is

$$\Delta(K) = K, \quad \Delta(1 \cdot K) = 1 \cdot K,$$

where $K$ is the Kolakoski word [10], whose first terms are

$$K = 22112122122112112212112122112112122122112122121121122\cdots$$

Clearly $K \in \mathcal{K}$, and we have $\Phi(K) = 2^\omega$ and $\Phi(1 \cdot K) = 1^\omega$.

The bijection $\Phi$ appears in the thesis of Lamas [12] and is used for a classification of infinite words. Independently, Dekking [8] used this bijection in order to show, for all $n \in \mathbb{N}$, the existence of words satisfying $\Delta^n(w) = w$. The Kolakoski word $K$ corresponds to the case $n = 1$.

It is easy to check that $\Delta$ commutes with the mirror image ($\sim$), is stable under complementation ($-$) and preserves palindromes:

**Proposition 2.** *For all $u \in \Sigma^*$ and for all $p \in \mathrm{Pal}(\Sigma^*)$ the operator $\Delta$ satisfies the conditions*

(a) $\Delta(\widetilde{u}) = \widetilde{\Delta(u)}$;
(b) $\Delta(\overline{u}) = \Delta(u)$;
(c) $\Delta(p) \in \mathrm{Pal}(\Sigma^*)$.

The following closure properties follow:

$$u \in \Delta_\Sigma^k \iff \overline{u},\ \widetilde{u} \in \Delta_\Sigma^k \quad \forall k \geqslant 0, \tag{7}$$

$$u \in \mathcal{K} \iff \overline{u} \in \mathcal{K}. \tag{8}$$

The fact that $\widetilde{u}$ does not appear in statement (8) is not surprising because closure by mirror image clearly involves two-sided infinite words, which are not considered here.

## 4. Avoidable and unavoidable patterns

First, the class $\mathcal{K}$ does not contain periodic words. Indeed, an eventually periodic word $W \in \mathcal{K}$ can always be written as $W = xu^\omega$, where $u$ is the smallest period also satisfying $\mathrm{Last}(x) = \mathrm{Last}(u) \neq \mathrm{First}(u)$, possibly by shifting conveniently the period. Then $\Delta(W) = \Delta(x)\Delta(u)^\omega$, and we have two cases. If $|\Delta(u)| = |u|$ then from conditions (2) and (3) we have $\Delta(W) = \Delta(x)1^\omega$. Otherwise, $\Delta(u)$ is a strictly smaller period, and an inductive argument establishes the claim (see also [6,12]).

In the case of the set of factors of $\mathcal{K}$, $\Delta$ is also a strict contraction except for a finite number of very small factors. For later use we quote this property in the following lemma.

**Lemma 3.** *For every $u \in F(\mathcal{K})$, such that $|u| > 4$ we have $|\Delta(u)| < |u|$.*

**Proof.** From condition (3) it suffices to show that $u$ does not contain 21212 or 12121. Suppose $u = x21212y$, then, $\Delta(u) = \Delta(x2).111.\Delta(2y)$, which implies that $u \notin F(\mathcal{K})$. The other case is similar. Consequently every factor with length $|u| > 4$ contains necessarily a block of length 2, i.e. the block 11 or/and the block 22. □

Every finite word $w \in \Delta_\Sigma^+$ can be easily extended to the right in a smooth word by means of the function $\Phi$ defined by (5):

$$\forall u \in \Sigma^\infty, \quad w \in \mathrm{Pref}(\Phi^{-1}(\Phi(w) \cdot u)).$$

Since an infinite smooth right extension $W$ of a smooth finite word $w$ contains $w$ as a factor, this means that the factors of the class $\mathcal{K}$ are *differentiable* in the sense of [3,7,20]. The left extensions require more work.

**Proposition 4.** *For all $v \in \Delta_\Sigma^k$, there exists $u \in \Delta_\Sigma^k$ such that $uv \in \Delta_\Sigma^{k+2}$.*

**Proof.** By definition $\Phi(v)$ ends with a 2, and consequently $\overline{\Phi(v)}$ ends with a 1. Then, compute $u' = \Phi^{-1}(\overline{\Phi(v)})$, and form the sequence of words,

$$w_n = \Delta^n\left(\widetilde{u'} \cdot v\right) = \Delta^n\left(\widetilde{u'}\right) \cdot \Delta^n(v), \quad n = 0, \ldots, k,$$

where the searched word $u = \widetilde{u'}$. Clearly, $w_k = 12$, therefore $\Delta(w_k) = 11$ and $\Delta^2(w_k) = 2$. Then we have, $w_0 = uv$ and $\Phi(w_0) = w_0[1]\cdots w_k[1]\,1\,2$. □

**Example.** Let $v = 21122122$, then $w_0 = 11212211 \cdot 21122122$, and $\Phi(w_0) = 121112112$,

$$
\begin{array}{cccccccccccccccc}
1 & 1 & 2 & 1 & 2 & 2 & 1 & 1 & \cdot & \mathbf{2} & \mathbf{1} & \mathbf{1} & \mathbf{2} & \mathbf{2} & \mathbf{1} & \mathbf{2} & \mathbf{2} \\
 & & 2 & 1 & 1 & 2 & 2 & \cdot & & \mathbf{1} & 2 & 2 & 1 & 2 & & & \\
 & & & & 1 & 2 & 2 & \cdot & & \mathbf{1} & 2 & 1 & 1 & & & & \\
 & & & & & 1 & 2 & \cdot & & \mathbf{1} & 1 & 2 & & & & & \\
 & & & & & 1 & 1 & \cdot & & \mathbf{2} & 1 & & & & & & \\
 & & & & & & 2 & \cdot & & \mathbf{1} & 1 & & & & & & \\
 & & & & & & 1 & \cdot & & \mathbf{2} & & & & & & & \\
 & & & & & & 1 & \cdot & & 1 & & & & & & & \\
 & & & & & & & & & 2 & & & & & & &
\end{array}
$$

The situation is different in the infinite case. The fact that $K$ and $1K$ are fixpoints of $\Delta$ implies that $K$ is not the proper suffix of any smooth word in $\mathcal{K}$, excepted $1K$.

**Lemma 5.** *For all $p \in \Sigma^+$ such that $|p| \geqslant 2$, we have $pK \notin \mathcal{K}$.*

**Proof.** Since smooth words do not contain 111 as a factor we may assume that $p$ ends with a 1. We have then $\Delta(pK) = \Delta(p)\Delta(K) = \Delta(p)K$. But by iterating, $\Delta^k(p) = 2$ for some $k$ so that $\Delta^k(pK) = 2 \cdot (2211\cdots)$. Therefore, $\Delta^{k+1}(pK)$ starts with a 3 which concludes the proof. □

We say that an infinite smooth word $W$ is *left extendable* if there exists an infinite smooth word $W'$ having $W$ as a proper suffix. For instance, $1K$ is not left extendable but $K$ is a proper suffix of only one word, namely $1K$. The next proposition gives a characterization of left extendable words in $\mathcal{K}$.

**Proposition 6.** *An infinite word $W \in \mathcal{K}$ is left extendable if and only if $\Phi(W) = u \cdot 2^\omega$, for some $u \neq 1$.*

**Proof.** ($\Longleftarrow$) If $u = \varepsilon$ the precedent Lemma 5 applies: $1K$ is the unique extension of $K$ and does not have an extension. If $|u| = k > 1$, we may assume that $u$ ends with 1 by removing the trailing 2's. Define

$$w = \Phi^{-1}(\overline{u}) = \Delta^{-k+1}_{\overline{u}[1..k-1]}(2),$$

so that we obtain

$$\Delta^k(\widetilde{w} \cdot W) = 1 \cdot K. \tag{9}$$

For instance, let $\Phi(W) = 211121 \cdot 2^\omega$, then we have

$$
\begin{array}{cccccccccccccc}
1 & 1 & 2 & 1 & 2 & 2 & 1 & 1 & \cdot & \mathbf{2} & \mathbf{1} & \mathbf{1} & \mathbf{2} & \mathbf{2} & \mathbf{1} & \mathbf{2}\,\mathbf{2}\cdots \\
 & & 2 & 1 & 1 & 2 & 2 & \cdot & & \mathbf{1} & 2 & 2 & 1 & 2 & \cdots & \\
 & & & & 1 & 2 & 2 & \cdot & & \mathbf{1} & 2 & 1 & 1 & \cdots & & \\
 & & & & & 1 & 2 & \cdot & & \mathbf{1} & 1 & 2 & \cdots & & & \\
 & & & & & 1 & 1 & \cdot & & \mathbf{2} & 1 & \cdots & & & & \\
 & & & & & & 2 & \cdot & & \mathbf{1} & 1 & \cdots & & & & \\
 & & & & & & 1 & \cdot & & \mathbf{2} & \cdots & & & & & \\
 & & & & & & \downarrow & \cdot & & \downarrow & & & & & & \\
 & & & & & & 1^\omega & \cdot & & 2^\omega & & & & & &
\end{array}
$$

where $w = \Phi^{-1}(122212)$ and the top line is $\widetilde{w} \cdot W$.

($\Longrightarrow$) We proceed by contradiction. Assume that for all $u$, $\Phi(W) = u \cdot v \neq u \cdot 2^\omega$. There are several cases to consider. If $u = 1^k$ and $v = 1^w$, then $W = 1K$ which is not left extendable. The case $u \neq 1^k$ and $v = 1^\omega$ corresponds to either

a finite word, which are not considered here, or to a word $W$ such that $\Delta^k(W) = 1K$, as shown in the example above, and Eq. (9). We may assume that $u$ ends with a 2, by putting the trailing 1's in the $v$. We also have

$$\Delta^{k-1}(W) = 2 \cdot \overline{K} = 2112212\cdots,$$

and the possible extensions are $2 \cdot 2112212\cdots$ which contains a cube, or $1 \cdot 2112212\cdots$ which is not smooth since $\Delta(1 \cdot 2112212\cdots) = 112211\cdots$.

For the last case, there must be an infinite number of occurrences of the factor 12 (and also 12) in $uv$. Therefore, assume that $uv = u' \cdot 12 \cdot v'$ for some $v' \notin \{1^\omega, 2^\omega\}$. Any finite extension $w \cdot W$ of $W$ must satisfy for some $k \geqslant 0$, $\Delta^k(wW) = \Delta^k(w) \cdot \Delta^k(W)$, with $\Delta^k(w) = 1$ and $\Delta^k(W) = 2\cdots$. Since there are infinitely many occurrences of 12 in $uv$, there exist an index $i > 0$ such that $\Delta^{k+i}(w) = 1$ and $\Delta^{k+i}(W) = 11\cdots$, and the factor 111 appears. Contradiction. □

A consequence is that proper suffixes of the Kolakoski word are not smooth.

**Corollary 7.** *For all $p \in \mathrm{Pref}(K)$, $p \neq \varepsilon$, we have $p^{-1}K \notin \mathcal{K}$.*

**Proof.** Let $U$ be a proper suffix of $K$. Observe that $U \neq K$, since otherwise $K$ would be periodic. Then, assume that $U$ is smooth, and left extendable to $K$. Therefore $\Phi(U) = u \cdot 2^\omega$, for some $u$ such that $\mathrm{Last}(u) = 1$. Let $|u| = k$, so that $\Delta^{k-1}(U) = \overline{K}$. This would imply that $\Delta^{k-1}(K) = 22\overline{K} = 221122\cdots$, contradiction. □

### *4.1. Squares and overlaps*

Carpi [4,5] established that the square factors in $K$ have length 2, 4, 6, 18 and 54. In fact this is a property of the class $\mathcal{K}$. The proof follows his scheme, but is sufficiently different that we provide it for the sake of completeness.

Define the set $S$ to be the smallest set in $F(\mathcal{K})$ satisfying the following conditions:

$$\begin{aligned} F_{\leqslant 7}(\mathcal{K}) &\subseteq S,\\ x \in S &\Longrightarrow \overline{x} \in S,\\ xy \in S &\Longrightarrow yx \in S,\\ x \in S, |x| = 2k &\Longrightarrow \Delta^{-1}(x) \in S. \end{aligned}$$

Since the set $S$ is closed under complementation and conjugation, to establish that $S$ is finite, it suffices to consider only minimal words (according to the lexicographic order) of the conjugacy classes.

**Lemma 8.** *The set $S$ is finite.*

**Proof.** Only the even-length words of $S$ expand in new words of $S$ and $|x| \leqslant 4$ implies that $|\Delta^{-1}(x)| \leqslant 7$. Therefore, it suffices to consider the following conjugacy classes of minimal smooth words (of length 6)

$$[112112], [112122], [112212], [122122].$$

The classes $\Delta^{-1}([112122])$ and $\Delta^{-1}([112212])$ contain words of length 9, so that their expansion terminates. Consider now $\Delta^{-1}([112112]) = \{[11212212]\}$, which contains words of length 8. Going further we obtain

$$\begin{aligned} \Delta^{-1}([11212212]) &= \{[112112212212], [112112122122]\},\\ \Delta^{-1}([112112212212]) &= \{[112112212212112122], [\overline{112112212212112122}]\},\\ \Delta^{-1}([112112122122]) &= \{[112112212112122122], [\overline{112112212112122122}]\}. \end{aligned}$$

The length of words on the right-hand side is 18, with an equal number of 1's and 2's so that the length of the next round of $\Delta^{-1}$ will be 27, ending the expansion. The last case is similar and the lengths obtained are 10 and 15. □

We need the following technical lemma (see [4]).

**Lemma 9.** *For any smooth finite word of the form $xyx$, with $|y| \leqslant 5$, one has $xy \in S$.*

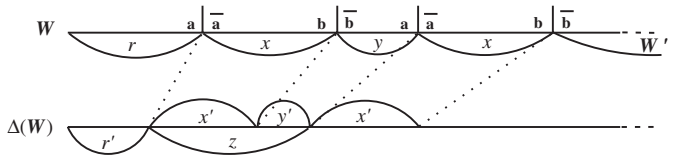


Fig. 1. The inductive step.

**Proof.** By definition, the statement is true when $|xy| \leqslant 7$. Let $W = rxyxW'$ be an infinite smooth word, and assume that $|xy| > 7$. By suitably factorizing one can also assume that

$$\mathrm{Last}(r) \neq \mathrm{First}(x) \neq \mathrm{Last}(xy) \quad \text{and} \quad \mathrm{Last}(x) \neq \mathrm{First}(yx) = \mathrm{First}(W'),$$

as shown in Fig. 1 (if $r$ is empty the conditions above still hold), where $a, b$ are letters in $\{1, 2\}$. Clearly, $|z| = \Delta(xy)$ is even and $\Delta(xyx) = x'y'x' \in F(\mathcal{K})$, with the conditions

$$|y'| \leqslant 5 \quad \text{and} \quad |x'y'| < |xy|.$$

The second condition is true by virtue of Lemma 3, and the inductive hypothesis. Finally, $z = x'y' \in S$ implies $\Delta^{-1}(z) \in S$ and therefore $xy \in S$ which completes the proof. □

A special case of Lemma 9 occurs when $|y| = 0$, that is, if $xx$ is smooth then $x \in S$, showing that $S$ contains the basis for squares. Since $S$ is finite, a systematic examination leads to the next proposition.

**Proposition 10.** *The lengths of smooth squares are* 2, 4, 6, 18 *and* 54.

Since each overlap $xuxux$ contains at least two squares, namely $xuxu$ and $uxux$, it follows that the number of overlaps is finite. More precisely, by a direct computation of smooth words one provides the following table where the factors numbered 3, 6, 8 and 13 are those given by Carpi in [4].

```
1-  2 12 2 12 2
2-  1 22 1 22 1
3-  21 2 21 2 21
4-  2 11 2 11 2
5-  1 12 1 12 1
6-  21 1 21 1 21
7-  1 21 1 21 1
8-  12 1 12 1 12
9-  1 22122112 1 22122112 1
10- 2 12211211 2 12211211 2
11- 2 11211221 2 11211221 2
12- 2 21 2 21 2
13- 12 2 12 2 12
14- 1 21122122 1 21122122 1
15- 1 211212211211221211211221221 2 1121212211211221211211221221 1
16- 2 122111211221211211221221211 2 1221112112212112112212212112 2
17- 1 211221221112122122112112122 1 2112212211212212211211212212 1
18- 2 122121122122112122122112112 2 1221211221221121221221121121 2
19- 2 112122122112112122112112212 2 1121221221121121221121121221 2
20- 2 112112212212112212211212212 2 1121122122121122122112112212 2
21- 1 221211211221221211221221121 1 2212112112212212112212212112 1
22- 1 221221121121221121121212112 1 2212211211212211211211212112 1
```

All the square factors but 22 and 11 are deduced from the overlaps, and by systematic inspection of the squares, it follows that

**Corollary 11.** *Smooth words are cube-free.*

*4.2. Palindromic loci : OK, ce "r" s'avère nul; lune! rêva Srečko*

In [3], the authors gave a characterization of smooth palindromes. More precisely, by using the left palindromic closure of a word $p$–Lpc($p$) for short—i.e. the smallest palindrome having $p$ for suffix, the class of smooth palindromes is (see [3])

$$\mathrm{Pal}(\mathcal{K}) = \mathrm{Lpc}(\mathrm{Pref}(\Delta^{-1}(1\cdot K)\cup\Delta^{-2}(1\cdot K))). \tag{10}$$

We take now a closer look to the occurrences of palindromes. The word $K$ does not have arbitrarily long palindromic prefixes. In fact, the only palindromic prefixes of $K$ are

**Lemma 12.** $\mathrm{Pal}(\mathcal{K})\cap\mathrm{Pref}(K) = \{\varepsilon, 2, 22\}$.

**Proof.** Any palindromic prefix $q$ of $K$ may be written as $q = p\cdot x\cdot\tilde{p}$ with $x\in\{\varepsilon,1,2\}$. When $|p|\leqslant 2$ this leads to the palindromic prefixes $\varepsilon, 2, 22$ of $K$. If $|p|>2$, then suppose first that $x\in\{1,2\}$. Then $\Delta(q) = p'\cdot 1\cdot\widetilde{p'}$ for some $p'\in\mathrm{Pref}(K)$. As $\Delta(q)\in\mathrm{Pref}(K)$, by the fixpoint property, we may repeat the argument until we get for some $k$, $\Delta^k(q) = 2\cdot 1\cdot 2$. Contradiction. If $x=\varepsilon$, then $q = p\tilde{p}$ and $p$ ends with 12 or 21. In both cases we have $\delta(q) = p'\cdot 2\cdot\widetilde{p'}$ for some $p'\in\mathrm{Pref}(K)$, reducing the problem to the previous case. □

As a consequence, we obtain that two occurrences of a smooth left palindromic closure of a prefix $q$ of $K$ are not too close.

**Proposition 13.** *Let $p,q\in\mathrm{Pref}(K)$ such that $|q|\leqslant|p|\leqslant 3|q|+1$. Then $x = \widetilde{q}\cdot 1\cdot q$ is a smooth palindrome and $p\cdot x\cdot\widetilde{p}\notin F(\mathcal{K})$.*

**Proof.** We have $p\cdot x\cdot\widetilde{p} = p\cdot(\widetilde{q}\cdot 1\cdot q)\cdot\widetilde{p} = (p\widetilde{q})\cdot 1\cdot(q\widetilde{p}) = (\widetilde{q\widetilde{p}})\cdot 1\cdot(q\widetilde{p})$, where the right-hand side is a palindrome if and only if $q\widetilde{p}\in\mathrm{Pref}(K)$. We have two cases. If $|p| = |q|$, the previous lemma applies. If $|p|>|q|$, then $p = qy$ and $q\widetilde{p} = q(\widetilde{qy}) = q\,\widetilde{y}\,\widetilde{q}$, and therefore $y$ is a palindrome of odd length. An inductive argument yields the result. □

On the other hand, there are words in $\mathcal{K}$ starting with arbitrarily long palindromes. Indeed, define the sequence of *full* prefixes of $K$ by

$$\mathrm{Full}(K) = \left\{ f_n = \Delta_2^{-n}(2^n)\mid n\in\mathbb{N}\right\}.$$

The first full prefixes of $K$ are

$$2, 22, 2211, 221121, 221121221, 22112122122112, \ldots$$

Every full prefix $f_n$ satisfies $\Delta(f_n) = f_{n-1}$ and can be written as

$$f_n = f_{n-1}\,u \quad \text{for some } u\in\Sigma^+.$$

Consequently,

$$w_n = \widetilde{f_n}\cdot 1\cdot\Delta_2^{-1}(f_n) = \widetilde{f_n}\cdot 1\cdot f_{n+1} = \widetilde{f_n}\cdot 1\cdot f_n\cdot u\in\Delta_\Sigma^*.$$

Equivalently, $w_n = \Delta_x^{-n}(2122)$ for a suitable $x\in\Sigma^n$, and we have the following result.

**Proposition 14.** *Let $w_n$ be defined as above. Then for each $n\in\mathbb{N}$, there exists an infinite set of words in $\mathcal{K}$ starting with $w_n$.*

**Proof.** For any word $U\in\Sigma^\omega$ we have $\Phi^{-1}(\Phi(w_n)\cdot U)\in\mathcal{K}$. □

As a direct consequence we have that every factor $u\in F(K)$ is also a factor of an infinite set of words $W\in\mathcal{K}-\{K\}$. Indeed it suffices to take a full prefix containing $u$.

**Corollary 15.** *For every factor* $u \in F(K)$, *we have* $\mathrm{Card}(\{W \in \mathcal{K} \mid u \in F(W)\}) = \infty$.

Moreover, with a bit of care, we can do better. According to Proposition 4 every full palindrome $Q$, can be extended to the left and to the right. Therefore, there exist infinite words containing almost arbitrary positions. All these results suggest, as supported by extensive computations, that all words in the class $\mathcal{K}$ share not only the same complexity but also the same factors, namely the factors of the Kolakoski word.

## 5. Fixed points of Δ and substitutions

Recall that $\Delta$ has two fixpoints, which are $\Delta(K) = K$ and $\Delta(1.K) = 1.K$. Furthermore, $\Phi(K) = 2^\omega$ and $\Phi(1.K) = 1^\omega$. It follows that $K$ is obtained as the fixpoint of the substitution (see [10,6])

$$\kappa : \begin{cases} 22 \to 2211 \\ 21 \to 221 \\ 12 \to 211 \\ 11 \to 21 \end{cases} \tag{11}$$

It is then easy to see that, for every integer $n$, $\Delta^n$ also has fixpoints (see [8,12]): each finite word $u$ of length $n$ satisfies

$$\Delta^n(\Phi^{-1}(u^\omega)) = \Phi^{-1}(u^\omega).$$

Moreover, for each $k$ such that $0 \leqslant k \leqslant n-1$ we have

$$\Delta^{n+k}(\Phi^{-1}(u^\omega)) = \Delta^k \Phi^{-1}(u^\omega),$$

so that all conjugates in the conjugacy class $[u]$ of $u$ also provide fixpoints. This is not surprising since the full shift $(\Sigma^*, s)$ ($s$ being the shift operator) is topologically conjugate of $(\mathcal{K}, \Delta)$ in the terminology of dynamical systems [13].

**Example.** For $n = 2$, there are 4 fixpoints for the operator $\Delta^2$ : we already know $\Phi(K) = (22)^\omega$ and $\Phi(1.K) = (11)^\omega$ which are also fixpoints for the operator $\Delta$; the other two are $\Phi(K_{12}) = (12)^\omega$ and $\Phi(K_{21}) = (21)^\omega$.

Now the question arises naturally whether there exist smooth words, besides $K$, that are obtained by some substitution. The answer is positive and relies on the existence of convenient codes. Recall that $X \subset F(\mathcal{K})$ is a code if every smooth word factorizes in at most one manner over $X$.

**Definition 16.** Let $C_n \subset F(\mathcal{K})$ be a code. $C_n$ is said to be *convenient* for $\Delta^n(W) = W$ if and only if

(i) $W = w_1 w_2 \cdots w_i \cdots$, with $w_i \in C_n$;
(ii) $\Delta_{1v}^{-k}(w_i) = 1x_i2$, and, $\Delta_{2v}^{-k}(w_i) = 2y_i1$, with $v \in \Sigma^{k-1}$, $1 \geqslant k \geqslant n$, for some smooth factors $x_i$, $y_i$.

Take for example $n = 1$. In order to have $\Delta_1^{-1}(w_i) = 1x_i2$ and $\Delta_2^{-1}(w_i) = 2y_i1$ for all words of $C_n$ we must take $w_i$ of even length. Thus, a convenient code is given by $C_1 = \{11, 12, 21, 22\}$. Of course $K$ or $1.K$ factorize over $C_1$ (because it contains all the words of length 2 in $F(K)$). Furthermore, we have

$$\Delta_2^{-1}(11) = 21,\ \Delta_2^{-1}(12) = 211,\ \Delta_2^{-1}(21) = 221,\ \Delta_2^{-1}(22) = 2211,$$
$$\Delta_1^{-1}(11) = 12,\ \Delta_1^{-1}(12) = 122,\ \Delta_1^{-1}(21) = 112,\ \Delta_1^{-1}(22) = 1122,$$

so that $C_1$ is a convenient code. Remark that the operator $\Delta_2^{-1}$ applied to $C_1$ is exactly the well-known substitution $\kappa$ given above (11) that generates $K$, while $\Delta_1^{-1}$ applied to $C_1$ defines a substitution $\kappa'$ that generates $1.K$:

$$\kappa' : \begin{cases} 22 \to 1122 \\ 21 \to 112 \\ 12 \to 122 \\ 11 \to 12 \end{cases}$$

The definition of convenient code $C_n$ simply ensures that, for $1 \leqslant k \leqslant n$, applying a sequence of $\Delta_1^{-1}$ or $\Delta_2^{-1}$ to every code word produces at each step a word starting and ending with complement letters $a$ and $\bar{a}$.

**Proposition 17.** *The following set is a convenient code for* $\Delta^2(W) = W$,

$$C_2 = \{22, 11, 2112, 2121, 212212, 21221121, 1221, 1212, 121121, 12112212\}.$$

**Proof.** First, a direct examination shows that $C_2$ is a prefix code. Now, let $v$ be an element of $C_2$. Then by definition we have

$$\Delta_{11}^{-2}(v) = 1w2,\ \Delta_{12}^{-2}(v) = 1x2,\ \Delta_{22}^{-2}(v) = 2y1,\ \Delta_{21}^{-2}(v) = 2z1,$$

for some $w, x, y, z \in \Sigma^*$. By construction $\Delta_a^{-1}(v) = aq\bar{a}$ with $a \in \Sigma$ and if $v = v_1 v_2 \cdots v_n$ then $n$ must be even in order to have a word starting with the letter $a$ and ending with the letter $\bar{a}$. Now, $\Delta_{ab}^{-2}(v) = \Delta_a^{-1}(bx\bar{b}) = aw\bar{a}$, and by the same parity argument the word $bx\bar{b}$ must have even length. But $|bx\bar{b}| = |\Delta_b^{-1} v| = 2|v|_2 + |v|_1$ and then the number of 1's in $v$ must be even.

To summarize $C_2$ is a finite set of even-length words having an even number of 1's constructed as follows. First, 22 and 11 are of even length with an even number of 1's and therefore belong to $C_2$. The factors 21 and 12 are of even length but odd number of 1's. The only trouble is if 21 or 21 can be right extended by a sequence of words of even length and odd number of 1's. For example 2122, 212211 is the beginning of such a sequence. But the next step is 21221122 which is not in $F(\mathcal{K})$ because $\Delta(\Delta(21221122)) = \Delta(11222) = 23$. Thus, the set $C_2$ is finite. The factor 21 can be extended to give $\{2112, 2121, 212212, 21221121\}$ and the factor 12 by $\{1221, 1212, 121121, 12112212\}$. □

The substitution associated with the convenient code $C_2$ is

$$\begin{aligned}
\Delta_{21}^{-2}(22) &= \Delta_2^{-1}(1122) = 212211,\\
\Delta_{21}^{-2}(11) &= \Delta_2^{-1}(12) = 211,\\
\Delta_{21}^{-2}(2112) &= \Delta_2^{-1}(112122) = 212212211,\\
\Delta_{21}^{-2}(2121) &= \Delta_2^{-1}(112112) = 21221211,\\
\Delta_{21}^{-2}(212212) &= \Delta_2^{-1}(1121122122) = 212212112212211,\\
\Delta_{21}^{-2}(21221121) &= \Delta_2^{-1}(112112212112) = 21221211221221211,\\
\Delta_{21}^{-2}(1221) &= \Delta_2^{-1}(122112) = 211221211,\\
\Delta_{21}^{-2}(1212) &= \Delta_2^{-1}(122122) = 2112212211,\\
\Delta_{21}^{-2}(121121) &= \Delta_2^{-1}(12212112) = 211221221211,\\
\Delta_{21}^{-2}(12112212) &= \Delta_2^{-1}(122121122122) = 2112212212112212211.
\end{aligned}$$

This 10 rules substitution yields the fixpoint $\Delta_{21}^{-2} K_{21} = K_{21}$ factorizing over $C_2$ as

$$K_{21} = 212212 \cdot 11 \cdot 22 \cdot 1221 \cdot 121121 \cdot 22 \cdot 11 \cdot 2112 \cdots.$$

The whole construction of $\Phi(K_{21}) = (21)^\omega$ is summarized now

$$\begin{aligned}
\Delta_{21}^{-2}(K_{21}) = K_{21} &= 212212112212211 \cdot 211 \cdot 212211 \cdot 211221211 \cdot\\
&\qquad 211221221211 \cdot 212211 \cdot 211 \cdot 212212211 \cdots\\
\Delta_1^{-1}(K_{21}) = \Delta(K_{21}) &= 1121122122 \cdot 12 \cdot 1122 \cdot 122112 \cdot 12212112 \cdot 1122 \cdot 12 \cdot 112122 \cdots\\
K_{21} = \Delta^2(K_{21}) &= 212212 \cdot 11.22 \cdot 1221 \cdot 121121 \cdot 22 \cdot 11 \cdot 2112 \cdots
\end{aligned}$$

The substitution associated with the operator $\Delta_{12}^{-2}$ is obtained in a similar way and is left to the reader.

For $n \geqslant 3$, the same method can be applied but the problem remains to prove that the minimal set of words $C_n$ is finite. For example, the definition imposes that each word $w$ of $C_3$ is of even length, contains an even number of 1's and also has an even number of 1's in odd positions (consequently, an even number of 1's in even positions). Using these

facts the construction of a finite set of words is not guaranteed. In fact, showing that $C_n$ is finite for all $n$ is equivalent to show that the language of the smooth words is equal to the language of $K$, which is still a conjecture.

## 6. Uncited references

[9,14,17].

## Acknowledgements

The comments provided by the anonymous referees greatly improved the clarity and overall presentation of the paper. Part of this work was done in Firenze and Pinzano, where the first author was visiting Università di Firenze. Grazie Renzo and the Renzo connection: Elena, Elisa, Enrica, Nicola, e quelli di Siena, Simone, Andrea.

## References

[1] J. Berstel, Axel Thue's papers on repetition words: a translation, Publ. du LaCIM, 1995.
[2] S. Brlek, Enumeration of factors in the Thue–Morse word, Discrete Appl. Math. 24 (1989) 83–96.
[3] S. Brlek, A. Ladouceur, A note on differentiable palindromes, Theoret. Comput. Sci. 302 (2003) 167–178.
[4] A. Carpi, Repetitions in the Kolakovski sequence, Bull. EATCS 50 (1993) 194–196.
[5] A. Carpi, On repeated factors in $C^\infty$-words, Inform. Process. Lett. 52 (6) (1994) 289–294.
[6] F.M. Dekking, Regularity and irregularity of sequences generated by automata, Séminaire de théorie des nombres de Bordeaux, exposé 9, 1979–1980.
[7] F.M. Dekking, On the structure of self generating sequences, Sém. théorie des nombres de Bordeaux, exposé 31, 1980–1981.
[8] F.M. Dekking, What is the long range order in the Kolakoski sequence, in: R.V. Moody (Ed.), The Mathematics of Long-Range Aperiodic Order, Kluwer Academic Publishers, Dordrecht, 1997, pp. 115–125.
[9] X. Droubay, G. Pirillo, Palindromes and Sturmian words, Theoret. Comput. Sci. 223 (1999) 73–85.
[10] W. Kolakoski, Self generating runs, Problem 5304, Amer. Math. Monthly 72 (1965) 674;
Solution: Amer. Math. Monthly 73 (1966) 681–682.
[11] A. Ladouceur, Outil logiciel pour la combinatoire des mots, Mém. Maitrise en Math., UQAM, AC20U5511 M6258, 1999.
[12] P. Lamas, Contribution à l' étude de quelques mots infinis, Mém. Maitrise en Math., UQAM, AC20U5511 M4444, 1995.
[13] D. Lind, B. Marcus, Symbolic Dynamics and Coding, Cambridge University Press, Cambridge, 1995.
[14] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.
[15] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, Theoret. Comput. Sci. 183 (1997) 45–82.
[16] M. Morse, Symbolic dynamics, Amer. J. Math. 60 (1938) 815–866.
[17] G. Paun, How much Thue is Kolakovski?, Bull. EATCS 49 (1993) 183–185.
[18] A. Thue, Über unendliche Zeichenreihen, Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl., Christiana 7 (1906) 1–22.
[19] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl., Christiana 1 (1912) 1–67.
[20] W.D. Weakeley, On the number of $C^\infty$-words of each length, J. Combin. Theory Ser. A 51 (1989) 55–62.